

# 2.1

## Hypotheses and Sources of Data

### hypothesis

- a theory or statement that is either true or false

### statistics

- numerical data, or the collection, organization, and analysis of numerical data

Some people think that drivers of red cars are more likely than other drivers to have an accident. How can you tell if this opinion is valid? First, write a **hypothesis** that clearly states what you want to prove or disprove. You can use **statistics** to test whether the hypothesis is likely to be true.

For example, you could start with the hypothesis that the accident rate for red cars is higher than that for other cars. You could then use data from accident reports or insurance claims to see if your hypothesis is correct. Often, it is not practical to collect enough data to determine for certain whether a hypothesis is true.



## Investigate

### How do you state a hypothesis?

1. Write a hypothesis for each question.
  - a) What percent of students in your school will get a driver's licence this year?
  - b) Is television advertising more effective than newspaper advertising?
  - c) Do people consider price or brand more important when buying toothpaste?
  - d) Do boys and girls have different study habits?
  - e) How often do people use the phonebook?
  - f) Which sport do teenagers like the most?
2. Do all hypotheses have to include numbers? How could you classify different types of hypotheses?
3. **Reflect** What information would you need to test each hypothesis in step 1?

All hypotheses are either true or false. If a hypothesis is false, then its opposite must be true.

## Example 1 State a Hypothesis and Its Opposite

Write a hypothesis about a relationship between the variables in each pair. Then, give the opposite hypothesis.

- a) a driver's age and the risk of having an accident
- b) attendance at school and marks
- c) the heights of boys and the heights of girls

### Solution

There are many possible hypotheses. Here is one example for each pair of variables.

- a) *Hypothesis:* As drivers age, their risk of having an accident increases.  
*Opposite:* As drivers age, their risk of having an accident does not increase. You can state the opposite another way: As drivers age, their risk of having an accident either decreases or stays the same.
- b) *Hypothesis:* Students' marks decrease as they miss more classes.  
*Opposite:* Students' marks do not decrease as they miss more classes.
- c) *Hypothesis:* On average, boys and girls of the same age are the same height.  
*Opposite:* On average, boys and girls of the same age are not the same height.

## Data Sources

Often, you need data to test a hypothesis. Researchers must decide whether to collect new data or use data that other people have already collected. **Primary data** come from experiments and surveys done by the researchers. Researchers can find **secondary data** in sources such as publications, the Internet, and surveys done by Statistics Canada.

### primary data

- original data that a researcher gathers specifically for a particular experiment or survey

### secondary data

- data that someone else has already gathered for some other purpose

## Example 2 Identify Primary and Secondary Data

Explain whether each set of data is primary or secondary. What are the advantages and any disadvantages of each person's choice of data source?

- a) Daniel telephoned 100 families in his town to ask them how many pets they have.
- b) Cathy used data from Statistics Canada to determine the proportion of households in Canada that have at least one car.
- c) Anja found a Web site with the results from a survey on the spending habits of teenagers across Canada.
- d) Tomas checked the Web sites of 24 stores for the price of the latest Harry Potter DVD.

### Did You Know?

Your school can get free access to data from Statistics Canada through its educational Web resource,  $\Sigma$ -STAT.

### Solution

- a) The telephone interviews produce primary data because Daniel performed the survey himself. The telephone survey is easy to do, but time-consuming. Data on pets in his town are unlikely to be available from a secondary source.
- b) Cathy is using a secondary source since Statistics Canada gathered the survey data. Statistics Canada is an excellent source because it collects data from a huge number of families all across Canada. Cathy could never gather that much data by herself.
- c) The Web site is a secondary source since Anja did not collect the survey data herself. The Web site is a convenient way to get survey results. It might take Anja a lot of time and expense to gather similar data herself. However, data on Web sites are not always reliable. Anja should check who did the survey and whether the results are complete and accurate. When a business or organization does a survey, they sometimes publish only data that are favourable to them.
- d) Tomas used the Web sites as a way to survey the prices that the video stores charged for the DVD. So, Tomas collected primary data from the Internet. This method was probably faster and cheaper than phoning all the video stores.

### Key Concepts

- A hypothesis is a statement that can be tested to determine if it is likely to be true.
- Primary data are new information collected by a researcher.
- Secondary data are existing data that were gathered for some other purpose.

### Communicate Your Understanding

- C1** a) Give an example of a hypothesis that involves numbers.  
b) Give an example of a hypothesis that does not involve numbers.  
c) Describe how you could test each hypothesis.
- C2** Is it always better to use primary data when testing a hypothesis? Explain why or why not.

## Practise

For help with questions 1 and 2, see Example 1.

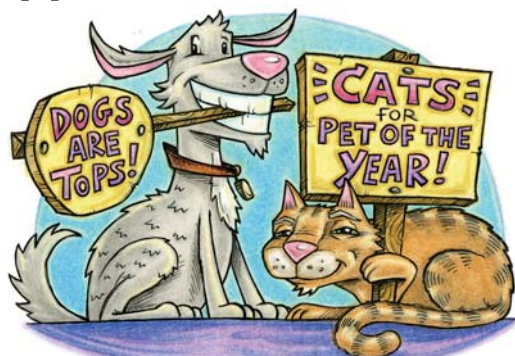
1. State the opposite of each hypothesis.
  - a) Most people's favourite number is 7.
  - b) Adults spend more time listening to classical music than to rap.
  - c) In Ontario, more teenagers join soccer teams than hockey teams.
  - d) Chocolate is not the most popular flavour of ice cream.
2. State a hypothesis about a relationship between each pair of variables. Then, state the opposite of each hypothesis.
  - a) a student's age and time spent doing homework
  - b) a mother's height and the height of her children
  - c) temperature and crime rates
  - d) the cost of gasoline and the number of people using public transit

For help with questions 3 and 4, see Example 2.

3. Which of the following data are primary and which are secondary? Explain.
  - a) An office manager hands out a questionnaire to see if employees want to work earlier hours during the summer.
  - b) A student finds data on Internet use in a report published by Statistics Canada.
  - c) A researcher collects information about how far people travel on public transit by talking to passengers on the buses.
  - d) A researcher downloads data about the length of rides taken on public transit from a transit authority's Web site.
4. Identify each data source as primary or secondary. State one advantage of each source of data.
  - a) A researcher interviewed 100 students about their study habits.
  - b) A sporting goods company searched on the Internet for data on how Canadians spend their leisure time.
  - c) A manufacturer surveyed 1000 recent customers about possible changes to a product.
  - d) A student found advertisements in out-of-town newspapers at a library to check admission prices at theatres across the country.

## Connect and Apply

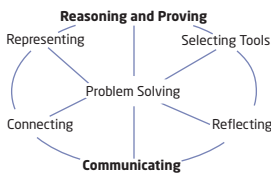
5.
  - a) Make a hypothesis about whether the students in your class prefer cats or dogs as pets.
  - b) Describe how you could test your hypothesis. Explain whether you would use primary or secondary data.



6. Steve prepared the following table using data volunteered by eight male students in his science class.

Name	Eye Colour	Height (cm)
Josanth	brown	167
Fred	green	181
Graham	green	185
Cho	brown	171
Seth	blue	154
Jamal	green	183
Juan	brown	160
Cameron	blue	173

- Is Steve using primary or secondary data? Explain.
- Make two hypotheses based on these data.
- How could you test your hypotheses?



- Make a hypothesis about the number of phone calls Canadians make.
  - Describe how you could use primary data to test your hypothesis.
  - Describe how you could use secondary data to test your hypothesis.
  - Which set of data is more likely to give accurate results?
- Chapter Problem** A coach is selecting students to compete in the high jump for the school's track and field team.
  - Make a hypothesis about a physical characteristic that could help an athlete do well in the high jump.
  - What data would you need to test your hypothesis? Would you use primary or secondary data? Explain why.
- Make a hypothesis about the relationship between the speed of a computer and its price.
  - Use Technology** Use an Internet search engine to collect data about computer prices. Compare the results when you use the key words "computer stores," "computer memory," and "computer prices."
  - Did you conduct primary or secondary research? Explain.
  - Describe another method for gathering data about computer speeds and prices.
- Use Technology**
  - Make a hypothesis about how much milk a cow produces in a day. Then, use an Internet search engine to find data to test your hypothesis.
  - Make a hypothesis about how much hay a cow eats in a day. Then, use data from the Internet to test your hypothesis.
  - Did you gather primary or secondary data in parts a) and b)? Explain.



## Achievement Check

11. This table shows the number of fish that entrants in a fishing derby caught.

Number of Fish	Number of Entrants
0	20
1	12
2	12
3	7
4	3
5	0
6	1

Before the fishing derby began, Heather predicted that most entrants would not catch any fish. George predicted that most entrants would catch either one or two fish.

- Is Heather's hypothesis true? Explain.
- Is George's hypothesis true? Explain.
- Write the opposite of Heather's hypothesis and of George's hypothesis.
- Can the opposite of Heather's hypothesis and the opposite of George's hypothesis both be true? Explain your reasoning.
- George found the results of the derby in a local newspaper. Are these data primary or secondary? Explain.
- Modify Heather's prediction and George's prediction to make new hypotheses that are true.



## Extend

- Make a hypothesis about the relationship between the latitude of a city and the mean of its daily maximum temperatures in January.
  - Use data from an atlas or an online source to test your hypothesis.
- Make a hypothesis about how the difference between the Olympic records for men and women in the marathon has changed over the years.
  - Use the Internet or other sources to collect data to test your hypothesis.
  - Explain how the data you found prove or disprove your hypothesis.
- Math Contest** The mean of a list of  $n$  numbers is 6. When the number 17 is added to the list, the mean becomes 7. What is the value of  $n$ ?

## Did You Know?

The coldest temperature ever recorded in North America was  $-63^{\circ}\text{C}$  at Snag, Yukon on February 3, 1947.